MINIMAL BACTERIAL GENOME

[0001] This application claims the benefit of the filing date of U.S. provisional application 60/725,295, filed Oct. 12, 2005, which is incorporated by reference herein in its entirety.

[0002] Aspects of this invention were made with government support (DOE grant number DE-FG02-02ER63453). The government has certain rights in the invention.

FIELD OF THE INVENTION

[0003] This invention relates, e.g., to the identification of non-essential genes of bacteria, and of a minimal set of genes required to support viability of a free-living organism.

BACKGROUND INFORMATION

[0004] One consequence of progress in the new field of synthetic biology is an emerging view of cells as assemblages of parts that can be put together to produce an organism with a desired phenotype (1). That perspective begs the question: "How few parts would it take to construct a cell?" In an environment that is free from stress and provides all necessary nutrients, what would comprise the simplest free-living organism? This problem has been approached theoretically and experimentally in our laboratory and elsewhere.

[0005] In a comparison of the first two bacterial genomes sequenced, Mushegian and Koonin projected that the 256 orthologous genes shared by the Gram negative *Haemophilus influenzae* and the Gram positive *M genitalium* genomes are a close approximation of a minimal gene set for bacterial life (2). More recently Gil et al. proposed a 206 protein-coding gene core of a minimal bacterial gene set based on analysis of several free-living and endosymbiotic bacterial genomes (3).

[0006] In 1999 some of the present inventors reported the first use of global transposon mutagenesis to experimentally determine the genes not essential for laboratory growth of *M genitalium* (4). Since then there have been numerous other experimental determinations of bacterial essential gene sets using our approach and other methods such as site directed gene knockouts and antisense RNA (5-12). Most of these studies were done with human pathogens, often with the aim of identifying essential genes that might be used as antibiotic targets. Almost all of these organisms contain relatively large genomes that include many paralogous gene families. Disruption or deletion of such genes shows they are non-essential but does not determine if their products perform essential biological functions. It is only through gene essentiality studies of bacteria that have near minimal genomes that we bring empirical verification to the compositions of hypothetical minimal gene sets.

[0007] The Mollicutes, generically known as the *mycoplasmas*, are an excellent experimental platform for experimentally defining a minimal gene set. These wall-less bacteria evolved from more conventional progenitors in the Firmicutes taxon by a process of massive genome reduction. Mycoplasmas are obligate parasites that live in relatively unchanging niches requiring little adaptive capability. *M genitalium* , a human urogenital pathogen, is the extreme manifestation of this genomic parsimony, having only 482 protein-coding genes and the smallest genome at ~580 kb of any known free-living organism capable of being grown in pure culture (13). The bacteria can grow independently on an agar plate free of other living cells. While more conventional bacteria with larger genomes used in gene essentiality studies have on average 26% of their genes in paralogous gene families, *M genitalium* has only 6% (Table 1). Thus, with its lack of genomic redundancy and contingencies for different environmental conditions, *M genitalium* is already close to being a minimal bacterial cell.

[0008] The 1999 report by some of the present inventors on the essential microbial gene for *M genitalium* and its closest relative, *Mycoplasma pneumoniae*, mapped ~2200 transposon insertion sites in these two species, and identified 130 putatively non-essential *M genitalium* protein-coding genes or *M pneumoniae* orthologs of *M genitalium* genes. In that report (Hutchison et al. (1999) *Science* 286, 2165-9), those authors estimated that 265 to 350 of the protein-coding genes of *M genitalium* are essential under laboratory growth conditions (4). However proof of gene dispensability requires isolation and characterization of pure clonal populations, which they did not do. In that report, the authors grew Tn4001 transformed cells in mixed pools for several weeks, and then isolated genomic DNA from those mixtures of mutants. They sequenced amplicons from inverse PCRs using that DNA as a template to identify the transposon insertion sites in the mycoplasma genomes. Most of the genes containing transposon insertions encoded either hypothetical proteins or other proteins not expected to be essential. Nonetheless, some of the putatively disrupted genes, such as isoleucyl and tyrosyl-tRNA synthetases (MG345 & MG455), DNA replication gene dnaA (MG469), and DNA polymerase III, subunit alpha (MG261) are thought to perform essential functions. They hypothesized how genes generally thought to be essential might be disrupted: a gene may be tolerant of the transposon insertion and not actually disrupted, cells could contain two copies of a gene, or the gene product may be supplied by other cells in the same mixed pool of mutants.

[0009] Disclosed herein is an expanded study in which we have isolated and characterized *M genitalium* Tn4001 insertion mutants that were present in individual colonies picked from agar plates. This analysis has provided a new, more thorough, estimate of the number of essential genes in this minimalist bacterium.

DESCRIPTION OF THE DRAWINGS

[0010] FIG. **1** shows the accumulation of new disrupted *M genitalium* genes (top line, thick) and new transposon insertion sites in the genome (bottom line, thin) as a function of the total number of analyzed primary colonies and subcolonies with insertion sites different from that of the parental primary colony.

[0011] FIGS. **2A-21** show global transposon mutagenesis of *M genitalium* . The locations of transposon insertions from the current study are noted by a Δ below the insertion site on the map. The letters over the Gene Loci (MG###) refer to the functional category of the gene product as listed.